Probability, Bayes, Midterm Review

1 Counting

1. Let p, q be prime. How many numbers are there among $1, 2, \ldots (pq)^2$ that are relatively prime to pq?

   Solution: This is an example of inclusion, exclusion. $p, 2p, \ldots, pqpq$ are not relatively prime to p, and $q, 2q, \ldots, pqpq$ are not relatively prime to q. That gives pq^2 numbers not relatively prime to p, and qp^2 not relatively prime to q. However, we have double counted the numbers that both p and q divide! These are $pq, 2pq, \ldots pqpq$, of which there are pq. So in total we have $pq^2 + qp^2 - pq$ numbers not relatively prime to p or q. But to find the numbers that are, we subtract that from all the numbers, $pqpq$, so we end up with $pqpq + pq - pq^2 - qp^2$.

2. Give a combinatorial proof of $\binom{k+n-1}{n-1} = \sum_{i=0}^{k} \binom{k-i+n-2}{n-2}$.

   Solution: On the left hand side, we can think of this as being k stars and n compartments, or equivalently k stars and $n-1$ bars. On the right hand side, we pick a certain number of stars to give to the first compartment, or equivalently to put left of the first bar. This number is represented by i. Then we have $n-2$ bars and $k-i$ stars remaining, which gives $\binom{k-i+n-2}{n-2}$ combinations.

2 Probability

1. Let A represent the event that someone has cancer. Let B be the event that they test for cancer and it comes positive. Assume that the probability that the test is positive is 90% if they have cancer and 10% if not, and that the overall percentage of people with cancer is 10%.

   (a) Write out all information we know in terms of $Pr()$ expressions.

   Solution:

   $$Pr(B|A) = 0.9$$
   $$Pr(B|\bar{A}) = 0.1$$
   $$Pr(A) = 0.1$$

   (b) What is the probability that someone has cancer if they tested positive?

   Solution: First draw a tree diagram to see what we don't know. Then, notice this is a great time to apply Bayes because we know more about $B|A$ than we do $A|B$. We are trying to find $Pr(A|B)$, but we can write:

   $$Pr(A|B) = Pr(A)\frac{Pr(B|A)}{Pr(B)} = 0.1\frac{0.9}{Pr(B)}$$

   We are left to find $Pr(B)$. But $Pr(B) = Pr(B|A)Pr(A) + Pr(B|\bar{A})Pr(\bar{A}) = 0.1 * 0.9 + 0.9 * 0.1 = 0.18$. Then we end up with 0.9/0.18, which is exactly half.

   (c) Say that the probability that someone dies if they have cancer is 2%, *and* this doesn't depend whether the test came back positive. Call the event that someone dies from cancer C, note that $P(C|\bar{A}) = 0$. Now say that the probability someone dies if they have surgery is p, but they are cured if they have surgery. At what p would it be a bad idea to have surgery if someone received a positive result on the test?

   Solution: It is a bad idea to have surgery if $p \geq P(C|B)$, i.e. if probability of death from surgery is at least the probability that you die if the test comes back positive. Then $P(C|B) = \frac{P(C\cap B)}{P(B)}$.

   But $P(C \cap B) = P(C \cap B|A)P(A) + P(C \cap B|\bar{A})P(\bar{A})$. It's impossible to die from cancer if you don't havec it, so then $P(C \cap B) = P(C \cap B|A) \cdot 0.1 = 0.9 * 0.2 * 0.1 = 0.0018$. Since $P(B) = 0.18$, we end up with $p \geq 0.01$ to be when we should not take surgery. This may defy your intuition, but remember that there's a half probability that someone doesn't have cancer even if the test says they do, and that in that case it's not worth to add risk of surgery.

2. Suppose $P(B|A) = P(B|\bar{A})$ where $\bar{A}$ is the complement of A. Prove that it must be the case that B is independent of A.

   Solution: To show that B is independent of A, we must show that

   $$P(B|A) = P(B)$$

But by the law of total probability,

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = P(B|A)P(A) + P(B|A)P(\bar{A})$$
$$= P(B|A)(P(A) + P(\bar{A})) = P(B|A)$$

We also use the fact that $P(A) + P(\bar{A}) = 1$.

3. True/False? For any events A,B,C in some probability space, $P(A \cap B \cap C) = P(A|B)P(B|C)P(C)$.

   Solution: False. Flip two coins, A represents two heads, B represents first one heads, C represents second one heads. $P(A \cap B \cap C) = P(A) = 1/4$, while the product is 1/8.

4. True/False? If events A and B are independent, so are $\bar{A}$ and B. Justify your answer with proof or counterexample.

   Solution: If they are independent, we have $P(A|B) = P(A)$. But then $P(\bar{A}|B) = 1 - P(A|B) = 1 - P(A) = P(\bar{A})$.

5. You flip a fair coin repeatedly. What is the probability that the first head you see is on the sixth flip?

   Solution: $1/2^6$. You have to see 5 tails and then a heads.

6. You flip a fair coin repeatedly. What is the probability that the second head you see is on the sixth flip?

   Solution: $\binom{5}{1}(1/2)^6$. You have to have a head somewhere in the first five and a head in the last 1 position. There are 5 such sample points and each has probability $(1/2)^6$.

7. Suppose 100 people stand in a line, in some random order, where Alice, Bob, and Chris are three of those people. If each permutation is equally likely, what is the probability that Bob is between Alice and Chris but not necessarily standing exactly next to them?

   Solution: This is 1/3. It doesn't matter how many people we are, but notice that we can just care about where Alice, Bob, and Chris are relative to each other. There are 3 permutations of Alice, Bob, and Chris by themselves, so this is 1/3.

8. Say we have a class of 30 people, and the probability that any two of them are friends on Facebook is 1/20. Give an upper bound for the probability that there are 5 students that are all friends with each other on Facebook.

   Solution: For 5 people to all know each other, we need $\binom{5}{2} = 10$ connections to happen, each of which happens independently. So the probability that any random set of 5 people is connected is around $\left(\frac{1}{20}\right)^5$. We can upper bound this by considering all possible $\binom{30}{5}$ combinations of 5 people, so the probability is $\leq \binom{30}{5}\left(\frac{1}{20}\right)^5 \approx 4.45\%$. For some good practice, figure out what the exact expression looks like.

3 Midterm Review

1. Alice wants to send Bob a message of n symbols (over $GF(p)$,where p is a prime) over a channel. The channel corrupts each symbol independently with probability q. Alice and Bob decide to use a Reed-Solomon code with Alice sending (n + m) symbols over the channel, and Bob using the Berlekamp-Welch decoding algorithm. If the probability that Bob cannot correctly decode Alice's message is to be kept at most α, then write an inequality (it can involve summations) that solves for the smallest value of m needed for this to be accomplished. (You can leave the equation in raw form but you must clearly express the dependencies on the parameters of the problem.)

   Solution: We want $P(\text{total packets corrupted} > m/2) \leq \alpha$.

   But $P(\text{total packets corrupted} > m/2) = \sum_{i>m/2}^{n+m} \binom{n+m}{i} q^i (1-q)^{n+m-i}$.

2. If I have a set T of k-bit strings, where $|T| = k$, give a procedure that looks at only one bit of each string and constructs a k-bit string that is not in the list. You can do things like let me look at the third bit of string 1, or the first bit of string 5.

   Solution: Let the ith bit of this bitstring be not the ith bit of the ith bitstring, then we are done.

3. Recall the following statement of the CRT: given k congruencies $x = a_i \pmod{n_i}$ where $a_i \neq 0$ and $gcd(n_i, n_j) = 1$ for $i \neq j$, there is exactly one $x \pmod{N}$ that satisfies all k congruencies for $N = \prod_i n_i$.

   (a) Consider that all the n_i are prime. Argue in this case that $x^{-1} \pmod{\prod_i n_i}$ exists.

   Solution: As x is relatively prime to all the n_is, it must be relatively prime with the product N too.

(b) Give an example where n_i may not be prime, where x does not have an inverse (mod N).

Solution: Where do we even start? Pick some small examples and enumerate! Take $n_1 = 4, n_2 = 3$, and see all the numbers that don't have inverses mod 12. The smallest one is 2, which satisfies our conditions since 2 is nonzero mod 4 and 3.

(c) Consider the case where every n_i is prime and we have $y = x^{-1} \pmod{N}$. What is $y \pmod{n_i}$? Justify this.

Solution: $y = a_i^{-1} \pmod{n_i}$. We know that $xx^{-1} = 1 + kN$, so looking mod n_i, we have that $a_i x^{-1} = 1 \bmod n_i$, and since all the a_is are nonzero and have unique inverses, we get our answer.

4. Consider the equation $(a_3x^3 + a_2x^2 + a_1x + a_0)(x_2 + b_1x + b_0) = F(x)(x_2 + b_1x + b_0)$, where $F(x)$ is some arbitrary function.

(a) Let $F(x) = a_3x^3 + a_2x^2 + a_1x + a_0$ on all but k points. What is the maximum value of k such that one can set b_1 and b_0 to "zero out" the equation?

Solution: $k = 2$, as we have a degree-2 polynomial we can set both roots of.

(b) For how many values of x do we need to know $F(x)$ to fully find a_3, a_2, a_1, a_0, b_1, and b_0, assuming that $F(x)$ differs from $a_3x^3 + a_2x^2 + a_1x + a_0$ on k points as found in the last problem?

Solution: This is exactly Berlekamp-Welch with $n = 4, k = 2$, so we need 8 points.

5. Consider the equation $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 70$, where each x_i is a non-negative integer.

(a) How many solutions to this equation are there?

Solution: $\binom{75}{5}$ by stars and bars. 70 stars, 5 bars.

(b) What if $x_1 \geq 30$ and $x_2 \geq 30$?

Solution: Give 30 each to both of them, now we have 10 balls and 6 bins so $\binom{15}{5}$.

(c) What if we require that either $x_1 \geq 30$ or $x_2 \geq 30$ or both?

Solution: Use inclusion/exclusion. If we require only one be at least 30, that gives $\binom{45}{5}$ possibilities remaining. We can add that twice, then subtract out $\binom{15}{5}$ to get rid of things that we double-counted.